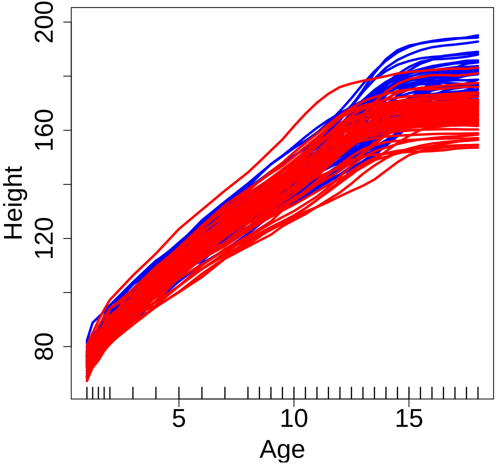
Functional Data Analysis in a Nutshell

Sarah Brockhaus
University of Mannheim, LMU Munich

15.02.2017

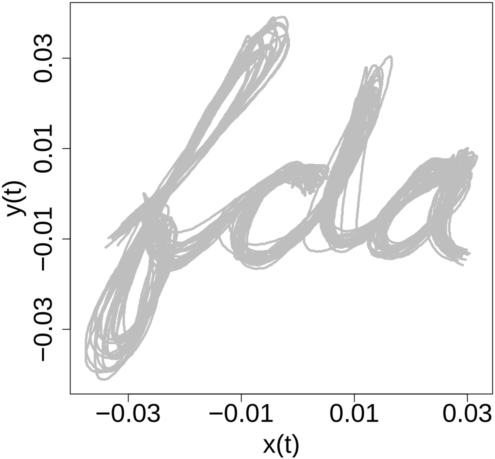
Introduction to functional data

Introduction to functional data



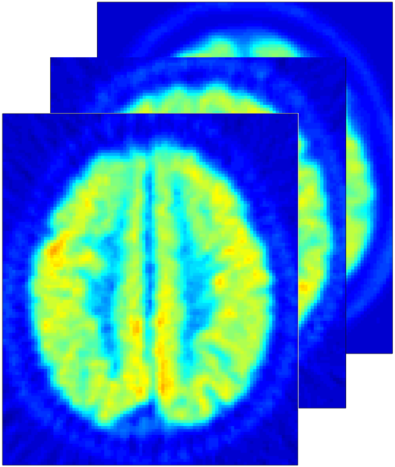
(Ramsay & Silverman, 2005)

Introduction to functional data

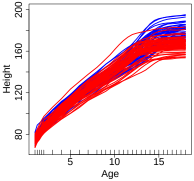


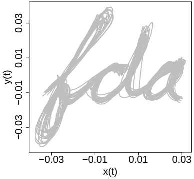
(Ramsay & Silverman, 2005)

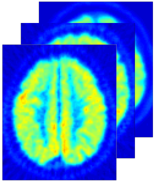
Introduction to functional data



Introduction to functional data

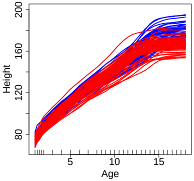


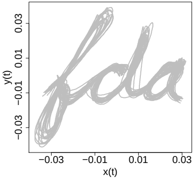


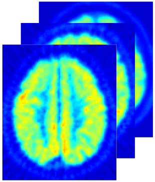


- Repeated measures for each observation unit; often measures over time

Introduction to functional data

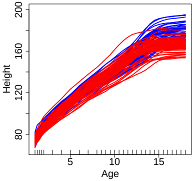


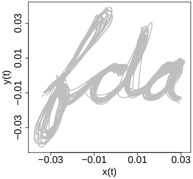


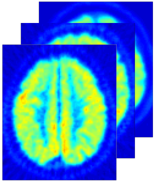


- Repeated measures for each observation unit; often measures over time
- Measures of irregular grids, „sparse data“

Introduction to functional data

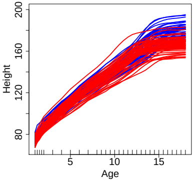


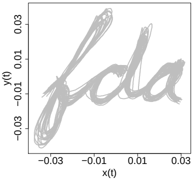


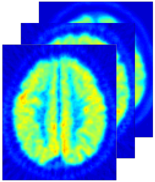


- Repeated measures for each observation unit; often measures over time
- Measures of irregular grids, „sparse data“
- Possibly arbitrary many measurements possible $\rightarrow$„smooth“ data generating function

Introduction to functional data

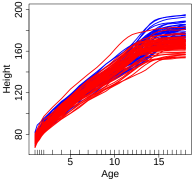


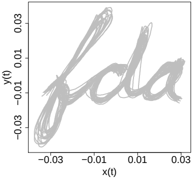


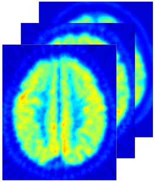


- Repeated measures for each observation unit; often measures over time
- Measures of irregular grids, „sparse data“
- Possibly arbitrary many measurements possible $\rightarrow$„smooth“ data generating function
- observations possibly with (measurement) error

Introduction to functional data







- Repeated measures for each observation unit; often measures over time
- Measures of irregular grids, „sparse data"
- Possibly arbitrary many measurements possible $\rightarrow$„smooth" data generating function
- observations possibly with (measurement) error
- many observations of the same data generating process and no prediction for further observations in the future $\leftrightarrow$ time series analysis

Progress of statistics

progress of mathematical statistics in terms of:

- sample space X: where the available data live
- parameter space Θ: where the target parameter belongs
- sample size n, number of variables d and number of parameters k

Statistical theory	**sample space X**	**parameter space Θ**
Classical parametric inference	$\mathbb{R}$	$\Theta \subset \mathbb{R}$
Multivariate analysis	$\mathbb{R}^d$ $(n \gg d)$	$\Theta \subset \mathbb{R}^k$ $(n \gg k)$
Nonparametrics	$\mathbb{R}^d$ $(n \gg d)$	a function space
High dimensional problems	$\mathbb{R}^d (n < d)$	$\Theta \subset \mathbb{R}^k$
Functional data analysis	a function space	$\mathbb{R}^k$ / a function space

(Cuevas, 2014)

Basic statistics for functional data

Mean, Variance and Covariance

- functional variable $X(t)$,
  with $t \in \mathcal{T}$ and $\mathcal{T}$ interval in $\mathbb{R}$
- sample $x_i(t)$, $i = 1, \ldots, n$

Mean, Variance and Covariance

- functional variable $X(t)$, with $t \in \mathcal{T}$ and $\mathcal{T}$ interval in $\mathbb{R}$
- sample $x_i(t)$, $i = 1, \ldots, n$
- functional mean:

$$\hat{\mu}_X(t) = \bar{x}(t) = \frac{1}{n} \sum_{i=1}^{n} x_i(t)$$

- functional variance:

$$\hat{\sigma}_X(t) = \frac{1}{n-1} \sum_{i=1}^{n} [x_i(t) - \bar{x}(t)]^2$$

Mean, Variance and Covariance

- functional variable $X(t)$, with $t \in \mathcal{T}$ and $\mathcal{T}$ interval in $\mathbb{R}$
- sample $x_i(t)$, $i = 1, \ldots, n$
- functional mean:

$$\hat{\mu}_X(t) = \bar{x}(t) = \frac{1}{n} \sum_{i=1}^{n} x_i(t)$$

- functional variance:

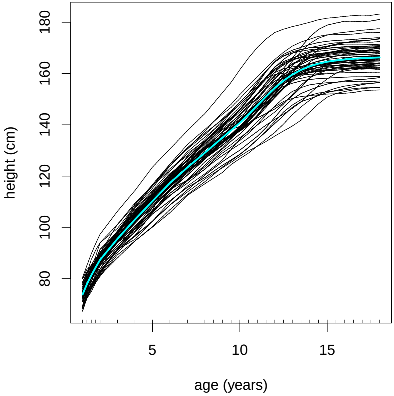
$$\hat{\sigma}_X(t) = \frac{1}{n-1} \sum_{i=1}^{n} [x_i(t) - \bar{x}(t)]^2$$

- functional (auto-)covariance:

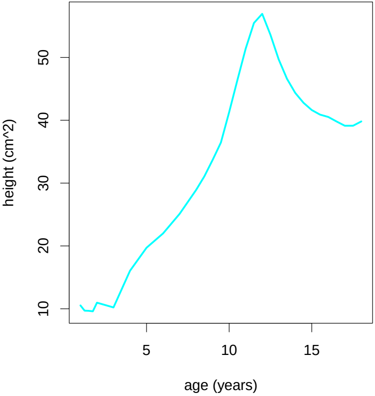
$$\hat{\sigma}_X(t_1, t_2) = \frac{1}{n-1} \sum_{i=1}^{n} [x_i(t_1) - \bar{x}(t_1)][x_i(t_2) - \bar{x}(t_2)]$$

Example for mean and variance

Growth of 54 girls

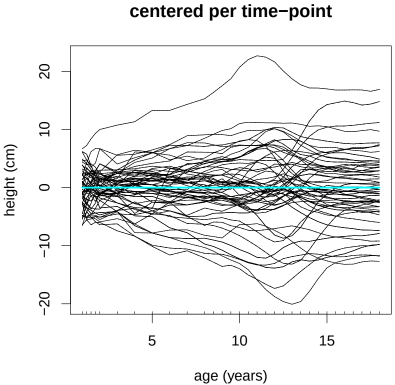


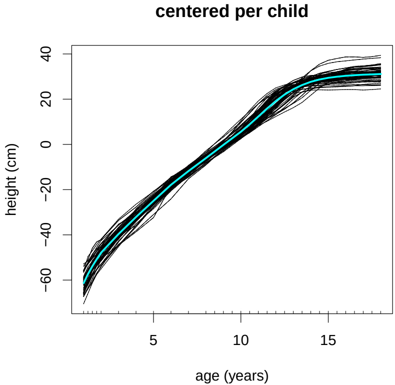
estimated mean



estimated variance

Two ways of centering data



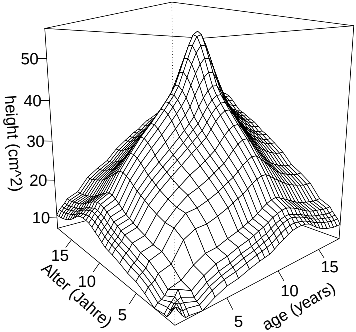


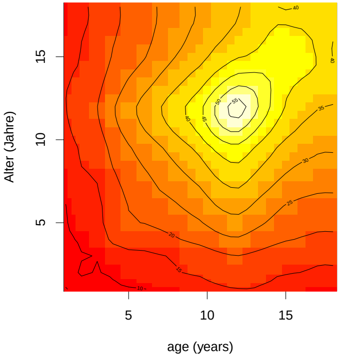
$$x_i^*(t) = x_i(t) - \bar{x}(t)$$

$$\tilde{x}_i(t) = x_i(t) - \int x_i(t)dt$$

Example for covariance surface

$$\hat{\sigma}_X(t_1, t_2) = \frac{1}{n-1} \sum_{i=1}^{n} [x_i(t_1) - \bar{x}(t_1)][x_i(t_2) - \bar{x}(t_2)]$$



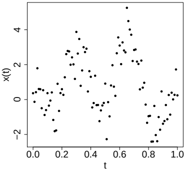


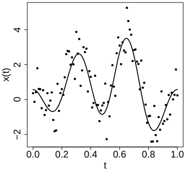
Outlook to functional data analysis

Outlook to functional data analysis

Important topics: (Ramsay & Silverman, 2005)

- Data representation → interpolation, smoothing

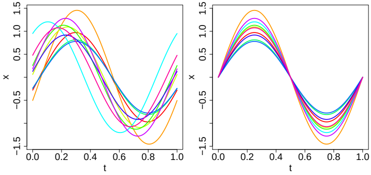


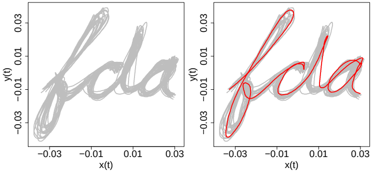


Outlook to functional data analysis

Important topics: (Ramsay & Silverman, 2005)

- Data representation → interpolation, smoothing
- Visualization → registration, outlyer detection

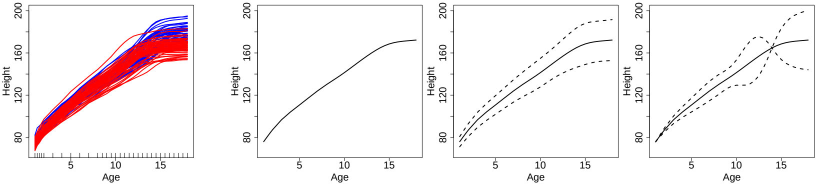




Outlook to functional data analysis

Important topics: (Ramsay & Silverman, 2005)

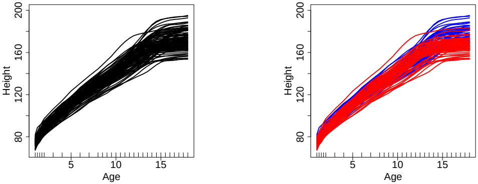
- Data representation $\rightarrow$ interpolation, smoothing
- Visualization $\rightarrow$ registration, outlyer detection
- Finding of patterns in the variation of the data $\rightarrow$ functional principal component analysis (FPCA)



Outlook to functional data analysis

Important topics: (Ramsay & Silverman, 2005)

- Data representation $\rightarrow$ interpolation, smoothing
- Visualization $\rightarrow$ registration, outlyer detection
- Finding of patterns in the variation of the data $\rightarrow$ functional principal component analysis (FPCA)
- Classification and clustering



Outlook to functional data analysis

Important topics: (Ramsay & Silverman, 2005)

- Data representation → interpolation, smoothing
- Visualization → registration, outlyer detection
- Finding of patterns in the variation of the data → functional principal component analysis (FPCA)
- Classification and clustering
- Regression → functional regression models (Greven and Scheipl, 2017)

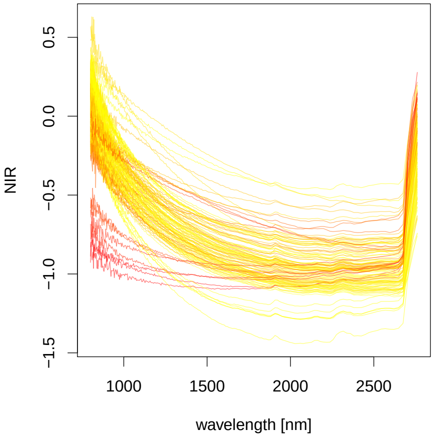
scalar-on-function: $$y_i = \mu + \int x_i(s)\beta(s)\mathrm{d}s + \varepsilon_i$$

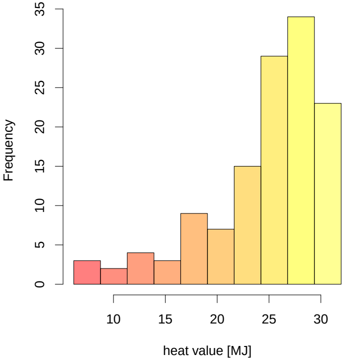
function-on-scalar: $$y_i(t) = \mu(t) + x_i\beta(t) + \varepsilon_i(t)$$

function-on-function: $$y_i(t) = \mu(t) + \int x_i(s)\beta(s,t)\mathrm{d}s + \varepsilon_i(t)$$

Spectral data of fossil fuels

Example for scalar-on-function-regression: Spectral data of fossil fuels to predict heat value

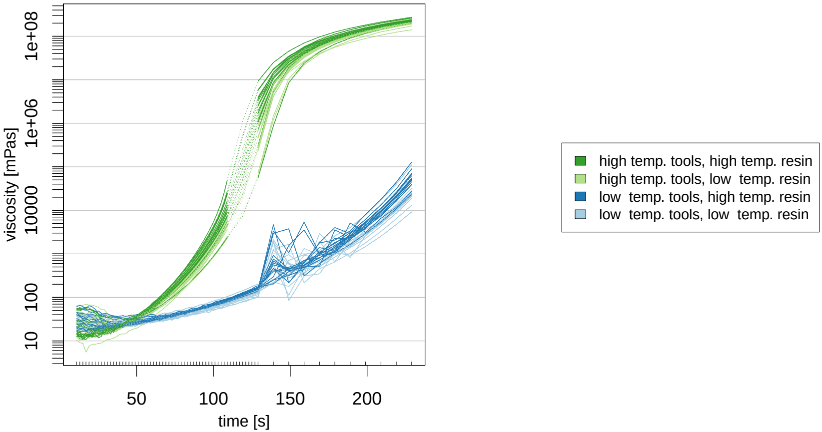




(Brockhaus et al. 2015)

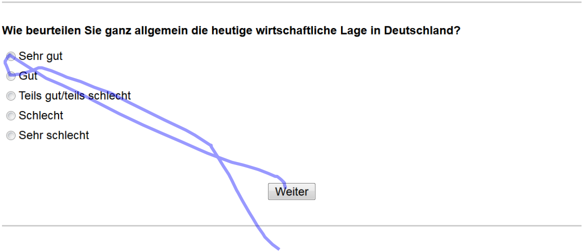
Viscosity of resin

Example for function-on-scalar-regression (functional ANOVA): Viscosity of resin depending on experimental conditions



(Brockhaus et al. 2015)

Work in progress: Analysis of mouse cursor movements



Literature and software

Literature and further reading

Monographs

- Ramsay & Silverman (2005), *Functional data analysis*, Springer, New York.
- Ferraty & Vieu (2006), *Nonparametric Functional Data Analysis*, Springer, New York

Overview articles

- Cuevas (2014), A partial overview of the theory of statistics with functional data, *Journal of Statistical Planning and Inference*, 147, 1–23.
- Levitin, Nuzzo, Vines & Ramsay (2007) Introduction to functional data analysis. *Canadian Psychology*, 48, 135–155.

Regression

- Greven & Scheipl (2017): A general framework for functional regression modelling. *Statistical Modelling*, to appear.
- Brockhaus, Scheipl, Hothorn & Greven (2015): The functional linear array model. *Statistical Modelling*, 15, 279–300.

R packages

Visualization

- Shang & Hyndman (2016). *rainbow: Rainbow Plots, Bagplots and Boxplots for Functional Data.* R package version 3.4. https://CRAN.R-project.org/package=rainbow

Visualization, descriptive and exploratory analysis

- Febrero-Bande & Oviedo de la Fuente (2012). *Statistical Computing in Functional Data Analysis: The R Package fda.usc.* Journal of Statistical Software, 51(4), 1–28.
- Ramsay, Wickham, Graves & Hooker (2014). *fda: Functional Data Analysis.* R package version 2.4.4. https://CRAN.R-project.org/package=fda

Regression

- Goldsmith, Scheipl, Huang, Wrobel, Gellar, Harezlak, McLean, Swihart, Xiao, Crainiceanu & Reiss (2016). *refund: Regression with Functional Data*. R package version 0.1-16. https://CRAN.R-project.org/package=refund

Functional Data Analysis in a Nutshell

Sarah Brockhaus

University of Mannheim, LMU Munich

15.02.2017